Office de la propriété intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An Agency of Industry Canada

Bureau canadien des brevets

Certification

La présente atteste que les documents ci-joints, dont la liste figure ci-dessous, sont des copies authentiques des documents déposés au Bureau des brevets.

Canadian Patent Office

Certification

This is to certify that the documents attached hereto and identified below are true copies of the documents on file in the Patent Office.

Specification and Drawings, as originally filed, with Application for Patent Serial No: **2,218,199** on December 9, 1997, by **MCGILL UNIVERSITY**, assignee of Guy A. Rouleau and Bernard Brais, for "Short GCG Expansions in the PAB II Gene for Oculopharyngeal Muscular Dystrophy and Diagnostic Thereof".

Tracy Faulkner
Agent certificateur/Certifying Officer

January 10, 2003
Date

Canada

(CIPO 68)
04-09-02

OPIC CIPO

ABSTRACT OF THE INVENTION

The present invention relates to a human PAB II gene containing transcribed polymorphic GCG repeat, which comprises a sequence as set forth in Fig. 4, which includes introns and flanking genomic sequence. The allelic variants of GCG repeat of the human PAB II gene are associated with a disease related with protein accumulation in nucleus, such as polyalanine accumulation, a disease related with swallowing difficulties, such as oculopharyngeal muscular dystrophy. The present invention also relates to a method for the diagnosis of a disease with protein accumulation in nucleus, which comprises the steps of: a) obtaining a nucleic acid sample of said patient; and b) determining allelic variants of GCG repeat of the gene of claim 1, and wherein long allelic variants are indicative of a disease related with protein accumulation in nucleus.

SHORT GCG EXPANSIONS IN THE *PAB II* GENE FOR OCULO-PHARYNGEAL MUSCULAR DYSTROPHY AND DIAGNOSTIC THEREOF

BACKGROUND OF THE INVENTION

(a) Field of the Invention

The invention relates to PAB II gene, and its uses thereof for the diagnosis, prognosis and treatment of a disease related with protein accumulation in nucleus, such as oculopharyngeal muscular dystrophy.

(b) Description of Prior Art

Autosomal dominant oculopharyngeal muscular dystrophy(OPMD) is an adult-onset disease with a world-wide distribution. It usually presents in the sixth decade with progressive swallowing difficulties (dysphagia), eye lid drooping (ptosis) and proximal limb weakness. Unique nuclear filament inclusions in skeletal muscle fibers are its pathological hallmark (Tome, F.M.S. & Fardeau, *Acta Neuropath.* **49**, 85-87 (1980)). We isolated the poly(A) binding protein II (PAB II) gene from a 217 kb candidate interval in chromosome 14q11. A (GCG)6 repeat encoding a polyalanine tract located at the N-terminus of the protein was expanded to (GCG)8-13 in the 144 OPMD families screened. More severe phenotypes were observed in compound heterozygotes for the (GCG)9 mutation and a (GCG)7 allele found in 2% of the population, whereas homozygosity for the (GCG)7 allele leads to autosomal recessive OPMD. Thus the (GCG)7 allele is an example of a polymorphism which can act as either a modifier of a dominant phenotype or as a recessive mutation. Pathological expansions of the polyalanine tract may cause mutated PAB II oligomers to accumulate as filament inclusions in nuclei.

It would be highly desirable to be provided with a tool for the diagnosis, prognosis and treatment

of a disease related with polyalanine accumulation in nucleus, such as oculopharyngeal muscular dystrophy.

SUMMARY OF THE INVENTION

One aim of the present invention is to provide a tool for the diagnosis, prognosis and treatment of a disease related with polyalanine accumulation in nucleus, such as oculopharyngeal muscular dystrophy.

In accordance with the present invention there is provided a human PAB II gene containing transcribed polymorphic GCG repeat, which comprises a sequence as set forth in Fig. 4, which includes introns and flanking genomic sequence.

The allelic variants of GCG repeat of the human PAB II gene are associated with a disease related with protein accumulation in nucleus, such as polyalanine accumulation, or with a disease related with swallowing difficulties, such as oculopharyngeal muscular dystrophy.

In accordance with the present invention there is also provided a method for the diagnosis of a disease with protein accumulation in nucleus, which comprises the steps of:

a) obtaining a nucleic acid sample of said patient; and

b) determining allelic variants of GCG repeat of the gene of the human PAB II gene, and wherein long allelic variants are indicative of a disease related with protein accumulation in nucleus, such as polyalanine accumulation and oculopharyngeal muscular dystrophy.

The long allelic variants have from about 245 to about 263 bp in length.

In accordance with the present invention there is also provided a non-human mammal model for the PAB II gene of the human PAB II gene, whose germ cells and

somatic cells are modified to express at least one allelic variant of the PAB II gene and wherein said allelic variant of the PAB II being introduced into the mammal, or an ancestor of the mammal, at an embryonic stage.

In accordance with the present invention there is also provided a method for the screening of therapeutic agents for the prevention and/or treatment of oculopharyngeal muscular dystrophy, which comprises the steps of:

a) administering said therapeutic agents to the non-human mammal of the present invention or oculopharyngeal muscular dystrophy patients; and

b) evaluating the prevention and/or treatment of development of oculopharyngeal muscular dystrophy in said mammal or said patients.

In accordance with the present invention there is also provided a method to identify genes part of or interacting with a biochemical pathway affected by PAB II gene, which comprises the steps of:

a) designing probes and/or primers using the hGT1 gene of the PAB II gene and screening oculopharyngeal muscular dystrophy patients samples with said probes and/or primers; and

b) evaluating the identified gene role in oculopharyngeal muscular dystrophy patients.

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A-B illustrate the positional cloning of the PAB II gene;

Figs. 2A-G illustrate the OPMD $(GCG)_n$ expansion sizes and sequence of mutations;

Fig. 3 illustrates the age distribution of swallowing time (st) for French Canadian OPMD carriers of the $(GCG)_9$ mutation; and

Fig. 4 illustrates the nucleotide sequence of human poly(A) binding protein II (hPAB II).

DETAILED DESCRIPTION OF THE INVENTION

In order to identify the gene mutated in OPMD, we constructed a 350 kb cosmid contig between flanking markers D14S990 and D14S1457 (Fig. 1A). Positions of the PAB II selected cDNA clones in relation to the EcoRI restriction map and the Genealogy-based Estimate of Historical Meiosis (GEHM)-derived candidate interval (Rommens, J.M. et al., in *Proceedings of the third international workshop on the identification of transcribed sequences* (eds. Hochgeschwender, U. & Gardiner, K.) 65-79 (Plenum, New York, 1994)).

The human poly(A) binding protein II gene (PAB II) is encoded by the nucleotide sequence as set forth in Fig. 4.

Twenty-five cDNAs were isolated by cDNA selection from the candidate interval (Rommens, J.M. et al., in *Proceedings of the third international workshop on the identification of transcribed sequences* (eds. Hochgeschwender, U. & Gardiner, K.) 65-79 (Plenum, New York, 1994)). Three of these hybridized to a common 20 kb EcoRI restriction fragment and showed high sequence homology to the bovine poly(A) binding protein II gene(bPAB II) (Fig. 1A). The PAB II gene appeared to be a good candidate for OPMD because it mapped to the genetically defined 0.26 cM candidate interval in 14q11 (Fig. 1A), its mRNA showed a high level of expression in skeletal muscle, and the PAB II protein is exclusively localized to the nucleus (Krause, S. et al., *Exp. Cell Res.* **214**, 75-82 (1994)) where it acts as a factor in mRNA polyadenylation (Whale, E., *Cell* **66**, 759-768 (1991); Whale, E. et al., *J. Biol. Chem.* **268**, 2937-2945 (1993); Bienroth, S. et al., EMBO J. **12**, 585-594 (1993)).

We subcloned a 8 kb HindIII genomic fragment containing the PAB II gene, and sequenced 6002 bp (GenBank: AF026029)(Nemeth, A. et al., *Nucleic Acids Res.* **23**, 4034-4041 (1995)) (Fig. 1B). Genomic structure of the PAB II gene, and position of the OPMD $(GCG)_n$ expansions. Exons are numbered. Introns 1 and 6 are variably present in 60% of cDNA clones. ORF, open reading frame; cen, centromere and tel, telomere.

The coding sequence was based on the previously published bovine sequence (GenBank: X89969) and the sequence of 31 human cDNAs and ESTs. The gene is composed of 7 exons and is transcribed in the cen-qter orientation (Fig. 1B). Multiple splice variants are found in ESTs and on Northern blots (Nemeth, A. et al., *Nucleic Acids Res.* **23**, 4034-4041 (1995)). In particular, introns 1 and 6 are present in more than 60% of clones (Fig. 1B)(Nemeth, A. et al., *Nucleic Acids Res.* **23**, 4034-4041 (1995)). The coding and protein sequences are highly conserved between human, bovine and mouse (GenBank: U93050). 93% of the PAB II sequence was readily amenable to RT-PCR- or genomic-SSCP screening. No mutations were uncovered using both techniques. However, a 400 bp region of exon 1 containing the start codon could not be readily amplified. This region is 80% GC rich. It includes a (GCG)6 repeat which codes for the first six alanines of a homopolymeric stretch of 10 (Fig. 2G). Nucleotide sequence of the mutated region of PAB II. Amino acid sequences of the N-terminus polyalanine stretch and position of the OPMD alanine insertions.

Special conditions were designed to amplify by PCR a 242 bp genomic fragment including this GCG-repeat. The (GCG)6 allele was found in 98% of French Canadian non-OPMD control chromosomes, whereas 2% of

chromosomes carried a (GCG)7 polymorphism (n=86) (Brais, B. et al., *Hum. Mol. Genet.* **4**, 429-434 (1995)).

Screening OPMD cases belonging to 144 families showed in all cases a PCR product larger by 6 to 21 bp than that found in controls (Fig. 2A). $(GCG)_6$ normal allele (N) and the six different $(GCG)_n$ expansions observed in 144 families.

Sequencing of these fragments revealed that the increased sizes were due to expansions of the GCG repeat (Fig. 2G). Fig. 2F shows the sequence of the (GCG)9 French Canadian expansion in a heterozygous parent and his homozygous child. Partial sequence of exon 1 in a normal $(GCG)_6$ control (N), a heterozygote (ht.) and a homozygote (hm.) for the $(GCG)_9$-repeat mutation. The number of families sharing the different (GCG)n-repeats expansions is shown in Table 1.

Table 1

Number of families sharing the different dominant $(GCG)_n$ OPMD mutations

Mutations	Polyalanine	Families
$(GCG)_8$	12	4
$(GCG)_9$	13	99
$(GCG)_{10}$	14	19
$(GCG)_{11}$	15	16
$(GCG)_{12}$	16	5
$(GCG)_{13}$	17	1
Total		**144**

**, 10 alanineresidues in normal PAB II.

The (GCG)9 expansion shared by 70 French Canadian families is the most frequent mutation we observed (Table 1). The (GCG)9 expansion is quite stable, with a single doubling observed in family F151 in an estimated 598 French Canadian meioses (Fig. 2C). The doubling of the French Canadian $(GCG)_9$ expansion is demonstrated in Family F151.

This contrasts with the unstable nature of previously described disease-causing triplet-repeats (Rosenberg, R.N., *New Eng. J. Med.* **335**, 1222-1224 (1996)).

Genotyping of all the participants in our clinical study of French Canadian OPMD provided molecular insights into the clinical variability observed in this condition. The genotypes for both copies of the PAB II mutated region were added to an anonymous version of our clinical database of 176 (GCG)9 mutation carriers (Brais, B. et al., *Hum. Mol. Genet.* **4**, 429-434 (1995)). Severity of the phenotype can be assessed by the swallowing time (st) in seconds taken to drink 80 cc of ice-cold water (Brais, B. et al., *Hum. Mol. Genet.* **4**, 429-434 (1995); Bouchard, J.-P. et al., *Can. J. Neurol. Sci.* **19**, 296-297 (1992)). The late onset and progressive nature of the muscular dystrophy is clearly illustrated in heterozygous carriers of the (GCG)9 mutation (bold curve in Fig. 3) when compared the average st of control (GCG)6 homozygous participants(n=76, thinner line in Fig. 3). The bold curve represents the average OPMD st for carriers of only one copy of the $(GCG)_9$ mutation (n=169), while the thinner line corresponds to the average st for $(GCG)_6$ homozygous normal controls(n=76). The black dot corresponds to the st value for individual VIII. Roman numerals refer to individual cases shown in Figs. 2B,2D and discussed in the text. Genotype of a homozygous $(GCG)_9$ case and her parents (Fig. 2B). Independent segregation of the $(GCG)_7$ allele. Case V has a more severe OPMD phenotype (Fig. 2D).

Two groups of genotypically distinct OPMD cases have more severe swallowing difficulties. Individuals I, II, and III have an early-onset disease and are

homozygous for the (GCG)9 expansion ($P < 10^{-5}$) (Figs. 2B, F). Cases IV, V, VI and VII have more severe phenotypes and are compound heterozygotes for the (GCG)9 mutation and the (GCG)7 polymorphism ($P < 10^{-5}$). In Fig. 2D the independent segregation of the two alleles is shown. Case V, who inherited the French Canadian (GCG)9 mutation and the (GCG)7 polymorphism, is more symptomatic than his brother VIII who carries the (GCG)9 mutation and a normal (GCG)6 allele (Figs. 2D and 3). The (GCG)7 polymorphism thus appears to be a modifier of severity of dominant OPMD. Furthermore, the (GCG)7 allele can act as a recessive mutation. This was documented in the French patient IX who inherited two copies of the (GCG)7 polymorphism and has a late-onset autosomal recessive form of OPMD (Fig. 2E). Case IX, who has a recessive form of OPMD, is shown to have inherited two copies of the (GCG)7 polymorphism.

This is the first description of short trinucleotide repeat expansions causing a human disease. The addition of only two GCG repeats is sufficient to cause dominant OPMD. OPMD expansions do not share the cardinal features of "dynamic mutations". The GCG expansions are not only short they are also meiotically quite stable. Furthermore, there is a clear cut-off between the normal and abnormal alleles, a single GCG expansion causing a recessive phenotype. The PAB II (GCG)7 allele is the first example of a relatively frequent allele which can act as either a modifier of a dominant phenotype or as a recessive mutation. This dosage effect is reminiscent of the one observed in a homozygote for two dominant synpolydactyly mutations. In this case, the patient had more severe deformities because she inherited two duplications causing an expansion in the polyalanine

tract of the HOXD13 protein (Akarsu, A.N. et al., *Hum. Mol. Genet.* **5**, 945-952 (1996)). A duplication causing a similar polyalanine expansion in the a subunit 1 gene of the core-binding transcription factor (CBFα1) has also been found to cause dominant cleidocranial dysplasia (Mundlos, S. et al., *Cell* **89**, 773-779 (1997)). The mutations in these two rare diseases are not triplet-repeats. The are duplications of "cryptic repeats" composed of mixed synonymous codons and are thought to result from unequal crossing over (Warren, S.T., *Science* **275**, 408-409 (1997)). In the case of OPMD, slippage during replication causing a reiteration of the GCG codon is a more likely mechanism (Wells, D.R., *J. Biol. Chem.* **271**, 2875-2878 (1996)).

Different observations converge to suggest that a gain of function of PAB II may cause the accumulation of nuclear filaments observed in OPMD (Tome, F.M.S. & Fardeau, *Acta Neuropath.* **49**, 85-87 (1980)). PAB II is found mostly in dimeric and oligomeric form (Nemeth, A. et al., *Nucleic Acids Res.* **23**, 4034-4041 (1995)). It is possible that the polyalanine tract plays a role in polymerization. Polyalanine stretches have been found in many other nuclear proteins such as the HOX proteins, but their functions is still unknown (Davies, S.W. et al., *Cell* **90**, 537-548 (1997)). Alanine is a highly hydrophobic amino acid present in the cores of proteins. In dragline spider silk, polyalanine stretches are thought to form B-sheet structures important in ensuring the fibers' strength (Simmons, A.H. et al., *Science* **271**, 84-87 (1996)). Polyalanine oligomers have also been shown to be extremely resistant to chemical denaturation and enzymatic degradation (Forood, B. et al., *Bioch. and Biophy. Res. Com.* **211**, 7-13 (1995)). One can speculate that PAB II oligomers comprised of a sufficient number of mutated

molecules might accumulate in the nuclei by forming undegradable polyalanine rich macromolecules. The rate of the accumulation would then depend on the ratio of mutated to non-mutated protein. The more severe phenotypes observed in homozygotes for the (GCG)9 mutations and compound heterozygotes for the (GCG)9 mutation and (GCG)7 allele may correspond to the fact that in these cases PAB II oligomers are composed only of mutated proteins. The ensuing faster filament accumulation could cause accelerated cell death. The recent description of nuclear filament inclusions in Huntington's disease, raises the possibility that "nuclear toxicity" caused by the accumulation of mutated homopolymeric domains is involved in the molecular pathophysiology of other triplet-repeat diseases (Davies, S.W. et al., *Cell* **90**, 537-548 (1997); Scherzinger, E. et al., *Cell* **90**, 549-558 (1997); DiFiglia, M. et al., *Science* **277**, 1990-1993 (1997)). Future immunocytochemical and expression studies will be able to test this pathophysiological hypothesis and provide some insight into why certain muscle groups are more affected while all tissues express PAB II.

Methods

Contig and cDNA selection

The cosmid contig was constructed by standard cosmid walking techniques using a gridded chromosome 14-specific cosmid library (Evans, G.A. et al., *Gene* **79**, 9-20 (1989)). The cDNA clones were isolated by cDNA selection as previously described (Rommens, J.M. et al., in *Proceedings of the third international workshop on the identification of transcribed sequences* (eds. Hochgeschwender, U. & Gardiner, K.) 65-79 (Plenum, New York, 1994)).

Cloning of the PAB II gene. Three cDNA clones corresponding to PAB II were sequenced (Sequenase,

USB). Clones were verified to map to cosmids by Southern hybridization. The 8 kb HindIII restriction fragment was subcloned from cosmid 166G8 into pBluescriptII (SK) (Stratagene). The clone was sequenced using primers derived from the bPABII gene and human EST sequences. Sequencing of the PAB II introns was done by primer walking.

PAB II mutation screening and sequencing. All cases were diagnosed as having OPMD on clinical grounds (Brais, B. et al., *Hum. Mol. Genet.* **4**, 429-434 (1995)). RT-PCR- and genomic SSCP analyses were done using standard protocols (Lafrenière, R.G. et al., *Nat. Genet.* **15**, 298-302 (1997)). The primers used to amplify the PAB II mutated region were: 5'-CGCAGTGCCCCGCCTTAGA-3' and 5'-ACAAGATGGCGCCGCCGCCCCGGC-3'. PCR reactions were performed in a total volume of 15 ml containing: 40 ng of genomic DNA; 1.5 mg of BSA; 1 mM of each primer; 250 mM dCTP and dTTP; 25 mM dATP; 125 mM of dGTP and 125 mM of 7-deaza-dGTP (Pharmacia); 7.5% DMSO; 3.75 mCi[35S]dATP, 1.5 unit of Taq DNA polymerase and 1.5 mM $MgCl_2$ (Perkin Elmer). For non-radioactive PCR reactions the [35S]dATP was replaced by 225 mM of dATP. The amplification procedure consisted of an initial denaturation step at 95°C for five minutes, followed by 35 cycles of denaturation at 95°C for 15 s, annealing at 70°C for 30 s, elongation at 74°C for 30 s and a final elongation at 74°C for 7 min. Samples were loaded on 5% polyacrylamide denaturing gels. Following electrophoresis, gels were dried and autoradiographs were obtained. Sizes of the inserts were determined by comparing to a standard M13 sequence (Sequenase, USB). Fragments used for sequencing were gel-purified. Sequencing of the mutated fragment using the Amplicycle kit (Perkin Elmer) was done with the 5'-

CGCAGTGCCCCGCCTTAGAGGTG-3' primer at an elongation temperature of 68°C.

Stability of (GCG)-repeat expansions. The meiotic stability of the (GCG)9-repeat was estimated based on our large French Canadian OPMD cohort. We previously established that a single ancestral OPMD carrier chromosome was introduced in the French Canadian population by three sisters in 1648. Seventy of the seventy one French Canadian OPMD families tested to date segregate a (GCG)9 expansion. However, in family F151, the affected brother and sister, despite sharing the French Canadian ancestral haplotype, carry a (GCG)12 expansion twice the size of the ancestral (GCG)9 mutation (Fig. 2C). In our founder effect study, we estimated that 450 (304-594) historical meioses shaped the 123 OPMD cases belonging to 42 of the 71 enrolled families. Our screening of our full set of participants allowed us to identify another 148 (GCG)9 carrier chromosomes. Therefore, we estimate that a single mutation of the (GCG)9 expansion has occurred in 598 (452-742) meioses.

Genotype-phenotype correlations. 176 carriers of at least one copy of the (GCG)9 mutation were examined during the early stage of the linkage study. All were asked to swallow 80 cc of ice-cold water as rapidly as possible. Testing was stopped after 60 seconds. The swallowing time (st) was validated as a sensitive test to identify OPMD cases (Brais, B. et al., *Hum. Mol. Genet.* **4**, 429-434 (1995); Bouchard, J.-P. et al., *Can. J. Neurol. Sci.* **19**, 296-297 (1992)). The st values for 76 (GCG)6 homozygotes normal controls is illustrated in Fig. 3. Analyses of variance were computed by two-way ANOVA (SYSTAT package). For the (GCG)9 homozygotes their mean st value was compared to the mean value for all (GCG)9 heterozygotes aged 35-40

($P < 10^{-5}$). For the (GCG)9 and (GCG)7 compound heterozygotes their mean st value was compared to the mean value for all (GCG)9 heterozygotes aged 45-65 ($P < 10^{-5}$).

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth, and as follows in the scope of the appended claims.

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. A human PAB II gene containing transcribed polymorphic GCG repeat, which comprises a sequence as set forth in Fig. 4, which includes introns and flanking genomic sequence.

2. The gene of claim 1, wherein allelic variants of GCG repeat are associated with a disease related with protein accumulation in nucleus.

3. The gene of claim 2, wherein said protein accumulation is polyalanine accumulation.

4. The gene of claim 1, wherein allelic variants of GCG repeat are associated with a disease related with swallowing difficulties.

5. The gene of claim 1, wherein said disease is oculopharyngeal muscular dystrophy.

6. A method for the diagnosis of a disease with protein accumulation in nucleus, which comprises the steps of:

a) obtaining a nucleic acid sample of said patient; and

b) determining allelic variants of GCG repeat of the gene of claim 1, and wherein long allelic variants are indicative of a disease related with protein accumulation in nucleus.

7. The method of claim 6, wherein said disease is oculopharyngeal muscular dystrophy.

8. The method of claim 7, wherein said long allelic variants have from about 245 to about 263 bp in length.

9. A non-human mammal model for the PAB II gene of claim 1, whose germ cells and somatic cells are modified to express at least one allelic variant of the PAB II gene and wherein said allelic variant of the PAB II being introduced into the mammal, or an ancestor of the mammal, at an embryonic stage.

10. A method for the screening of therapeutic agents for the prevention and/or treatment of oculopharyngeal muscular dystrophy, which comprises the steps of:

a) administering said therapeutic agents to the non-human mammal of claim 9 or oculopharyngeal muscular dystrophy patients; and
b) evaluating the prevention and/or treatment of development of oculopharyngeal muscular dystrophy in said mammal or said patients.

11. A method to identify genes part of or interacting with a biochemical pathway affected by PAB II gene, which comprises the steps of:

a) designing probes and/or primers using the hGT1 gene of claim 1 and screening oculopharyngeal muscular dystrophy patients samples with said probes and/or primers; and
b) evaluating the identified gene role in oculopharyngeal muscular dystrophy patients.

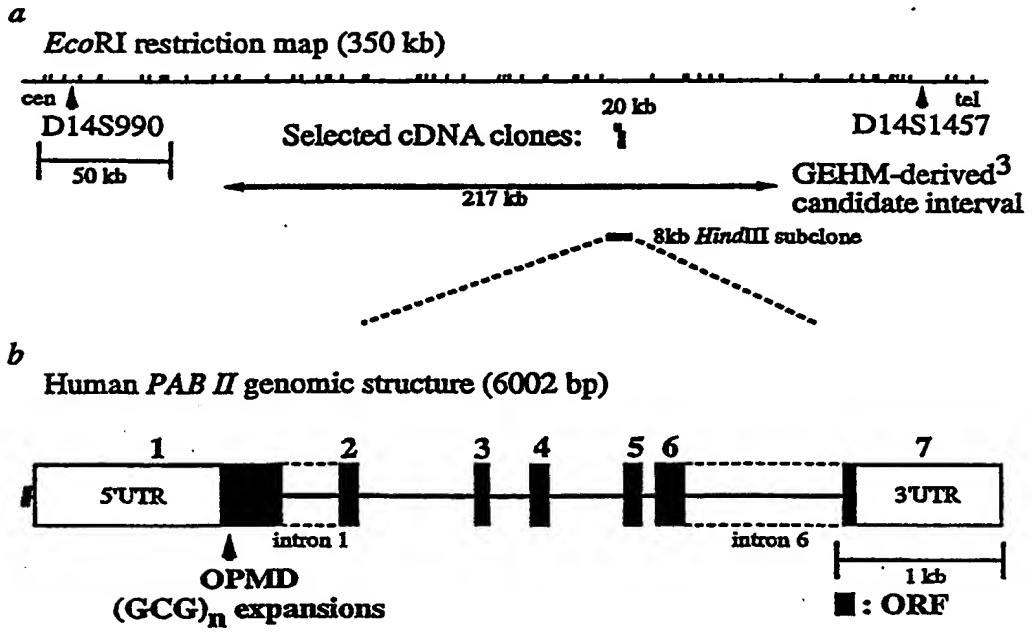

Fig. 1

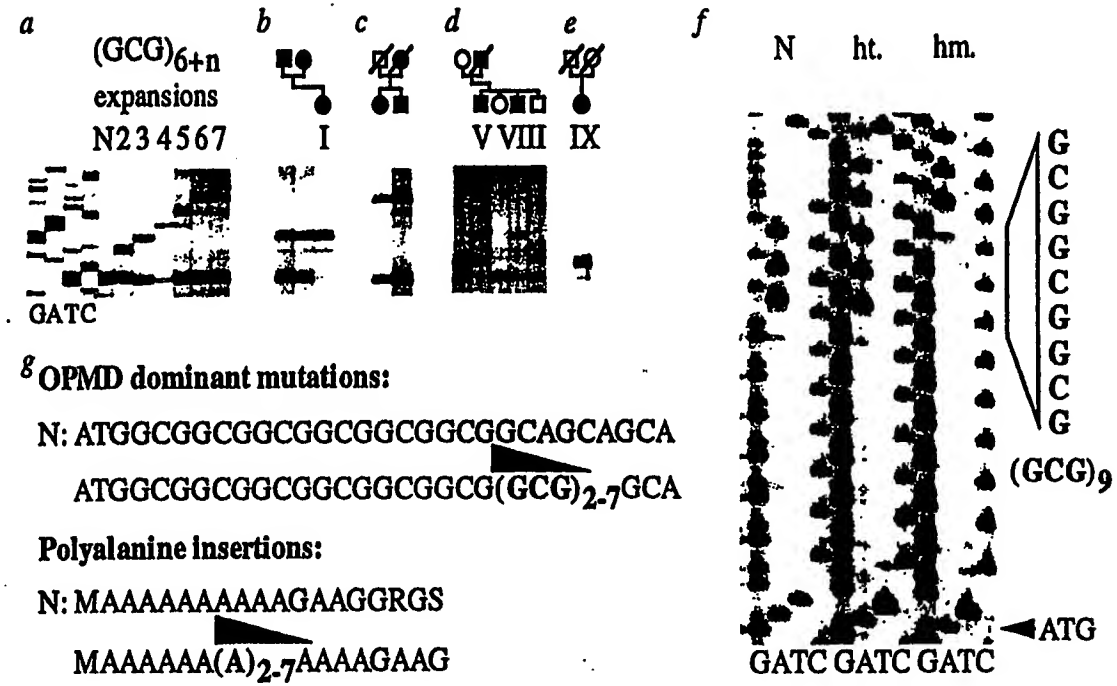

Fig. 2

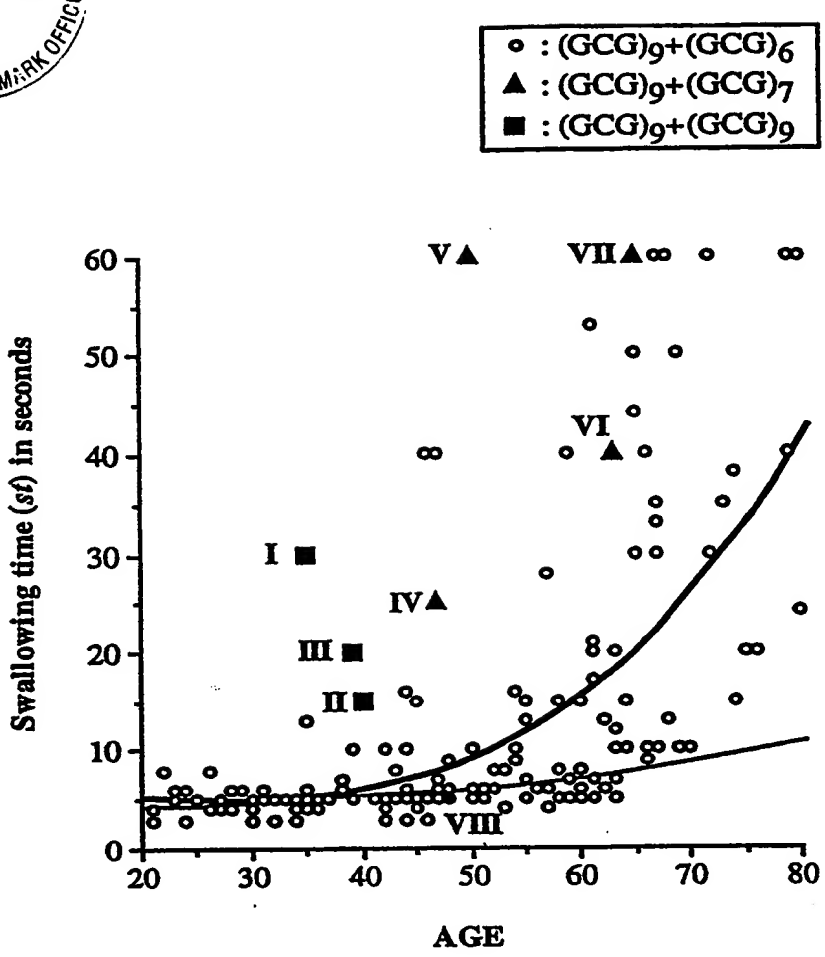

Fig. 3

```
   1 aatgaaggtg gacacccaaa tagccccaat acaaatgcct gttcaatcaa ccaaacatct
  61 aagcagcaca tctatgtggt agcatattgc caggccgtga gactgcgaat ataaatagga
 121 accgccccctc atctgcaggc gctcacaacc tagttagcaa acagtaaaac aattaagcgc
 181 gccgtggaca taggcccact tgtcctggga aatgagggga agctggggtt tgcagtggtt
 241 tgattgaagg gggactacat gttagaggca cagactgggt gcaggtacac ccaaaggaac
 301 gagaagagtg gaaggaaaca acatccacaa agtaaccaca tgctggcgta tcgaaggccg
 361 tgatttacgg ttttgagact ttacctcgcc agcaaagggg ggccagtctg ttagcggtgc
 421 agattggagg ggtgacattg gaagctgtcc aggaaaaaga aaatggaact ggggagcaga
 481 aggcctacgc aagagggcgg gacagacagg acttgtgact agtagctctg gactgaggaa
 541 tcctccctgc tttctggtgc gggagagcta gtggatgatg gtgccaataa cctggatggg
 601 gaaagtaagc tccctcctgg aatgcttcat tcacaacctc cattttcagc aacatcccat
 661 ctactggtgc ttcctggtcg agatacaagt ttcctgaaac tgctgctctg ttttgggcct
 721 cacccggcca acagctcact agctggcaag cagtagtatc aagatggcgg ccccctagga
 781 ctggctagtc atgtgacctc gggtttccca agtttgaagc ccggcagtcc tttcgggggc
 841 aaggttcacc tgtcacgaaa cgagtgtcac cccttcgact ctcgcaagcc aatcggcatc
 901 tgagactggg ccactgcggt gaggcgatcg gaagattggt cctttccagt cgcctagcta
 961 gggccaatca cggagcgtcc catacttcgc gggcccgccc gtaggccggg gagaagcagg
1021 aatatcgtca cagcgtggcg gtattattac ctaaggactc gataggaggt gggacgcgtg
1081 ttgattgaca ggcagatttc cctaccggga tttgagaatt tggcgcagtg cccgccttag
1141 aggtgcgctt atttgattgc caagtaatat tccccaatgg agtactagct catggtgacg
1201 ggcaggcagc ttgagctaat gagtcctccg tggccggcgc agctctccac atgccgggcg
1261 gcgggcccca gtctgagcgg cgatggcggc ggcggcggcg gcggcagcag cagcgggggc
1321 tgcgggcggt cggggctccg ggccggggcg gcggcgccat cttgtgcccg gggccggtgg
1381 ggaggccggg gaggggggcc cgggggcgc aggggactac gggaacggcc tggagtctga
1441 ggaactggag cctgaggagc tgctgctgga gcccgagccg gagcccgagc ccgaagagga
1501 gccgcccccgg ccccgcgccc ccccgggagc tccgggccct gggcctggtt cgggagcccc
1561 cggcagccaa gaggaggagg aggagccggg actggtcgag ggtgacccgg gggacggcgc
1621 cattgaggac ccggtgagga aggagggcga gcgagcaggc cggcggctgg cgcgtcactg
1681 gaggcccaga gctcgggcga gcggtggcag gcgggggtg gggttgggcg gggaataacg
1741 tggctggggc gggtcgggcc ggggatgggt cagcgatcac tacaaggggc ccgactggct
1801 tgattcgggc gtcacgggtg cctagtgttg ttctagagag ggtagctttt cttttatcac
1861 gaccctcgca tggggcgagg gaaatggccg agcatggctg aggcgcgctc tggccgagag
1921 cagggcacag cccctgcgtt ggttcctctt aagctgtcct ccataccctc cccacttata
1981 ttaggagctg gaagctatca aagctcgagt cagggagatg gaggaagaag ctgagaagct
2041 aaaggagcta cagaacgagg tagagaagca gatgaatatg agtccacctc caggcaatgc
2101 tgagtaactg gcggttgcac gcggagcccg ggttctcggg ttggaagggt tgtggggagg
2161 atggggaatg tggggttaga tactcggcac cctggagctg cttgtctgag ctattatgac
2221 tgtgccgcgg tcatagtccg ttgtgtgttc ctctgacctt tgtgaggcag aactgatatt
2281 ttggtggtgg tagccttgtg cctccctttg tcctgttata attgtgttgc tctttattct
2341 tagtctacgt ctatctttct ttggtagagg ttgcgtgctc gcatttgacc ttcaaatcta
2401 atagtttttc ctccaattgg agacgcttta ggattctaag agaaagcaag ctggaagggg
2461 tttccccttt aaattctaga aatgtggagt ctcagcccac ttaattttgc tcactcttaa
2521 aagcatttca accaaagcca ttcattaggg atttgatttg gagggcagga gggattccta
2581 tactgtttta agtgtgtatt aattctttca atttatcgaa ttatttagtg agtaacctgc
2641 tatgcactag gcactattct cggcttgtgg gtacagcagg gaacagcaca gaccaaaatc
2701 tttgccttca ctgagcttat gggatagtgc tggtggtgga agtgcaacat attggtcaag
2761 tagaaaacaa gtgtgtggtt tttgtaaaaa attatttttt cctgatagct ggcccggtga
2821 tcatgtccat tgaggagaag atggaggctg atgcccgttc catctatgtt ggcaatgtga
2881 cgtactgggg ctctgactgg ggttgggggc aagttcttct tttggggaat tatttaatag
2941 tcctgaaaga acatctccgg gatagatgtg gttttgggtg tggagggagt gtgggaagga
3001 ggttaaaggt aatggaatga tcagtaatca gcaaaggctc tgggtttgga aggaaaagag
3061 attaattcct caaattacca gatttcatgt gctttggtgt atgatggccc agaccaaagg
3121 ctcgggaggg ttcttttgag acaggaattt gcctggtgcc tgtgaaattt ttctcctctc
3181 atcaggtgga ctatggtgca acagcagaag agctggaagc tcactttcat ggctgtggtt
3241 cagtcaaccg tgttaccata ctgtgtgaca aatttagtgg ccatcccaaa ggtaaagtaa
3301 aggggagtaa gttgagataa tttaaattac agtgtacaaa tagataaatt atgttttata
3361 ttgagcagta agttatttgg tgttaacaca ggtgatctgt gtcatttaag atcatggcat
3421 taatgttgat atatcaggag ttgcacctaa atgtcttcag aggccagata acaaaaatga
3481 aggctagatg tgggtgggat tacgaactag aaggggaggg gcagcttcta cttggcctat
3541 tatggcatat ggaaattcag gccctgtgtg tcttattttt acaaatttca aagagtagct
3601 ggaaatttta aaatttaaat gatttcgaat gattgaaatt ttccatttag aagaattttg
3661 acaaataaaa aatataactg cattgtagcc caaaacgaag catgcctgca ggttgaattt
```

Fig. 4A

```
3721 gacctgtgag gtatttgtaa cctcagagag atacaatgac aattcttttc aggtttgcgt
3781 atatagagtt ctcagacaaa gagtcagtga ggacttcctt ggccttagat gagtccctat
3841 ttagaggaag gcaaatcaag gtaagcctat gtccattgct gttctagttg tgtataaact
3901 ctccaggttg cctttaaggc tatcatttgt tcatctctga ctcaggtgat cccaaaacga
3961 accaacagac caggcatcag cacaacagac cggggtttтc cacgagcccg ctaccgcgcc
4021 cggaccacca actacaacag ctcccgctct cgattctaca gtggttttaa cagcaggccc
4081 cggggtcgcg tctacaggtc aggatagatg ggctgctcct ctttcccccg cctcccgtga
4141 gccccgtatg cttcctcctc tctggtctga ggaacctccc tccccccacc cctccccgtg
4201 gtcttcagga actttgtctc ctgcctgtgc aggttgagga aggtagttgc aggccaggcc
4261 agaaggcagc ctcatcatct tttctgcagt agaaattggt gataagggct gcatccctcc
4321 cttggttcaa agaggcttcc acccccagcc ttttttttct tgggagttgg tggcatttga
4381 aggtgtttgc ggacaaaact gggaggaaca gggcctccag gaagttgaaa gcactgcttg
4441 gacatttgtt actttttтcg gagttaggga gggattgaag actgaacctc ccttggaaga
4501 ataccagagg ctagctagtt gatcctccca acagccttgt gggaggattt tgagatactt
4561 attctttatt tgagccagtc ttgcaaggtt aacttctcac tgggcctagt gtggtnccca
4621 ggtttttgcc ttgcttcact tctgtctcta catttaaata gacgggttag gcatataaac
4681 cttggctttt cataagctct acctgcctat ccccaggagt tagggaggat ctatttgtga
4741 aggccctagg gtttaaaaac tgtggaggac tgaaaaactg gataaaaagg gggtcctttt
4801 ccttgccccт gtctctcact cagatgcgct tctttttcgc cactgtttgg caaagttttc
4861 tgttaagccc ccctcccccт gccccagttc tcccaggtgc gttactattt ctgggatcat
4921 ggggtcggtt ttaggacact tgaacacttc ttttcccccc ttcccttcac agtaactggg
4981 gcaggggcct acggggaggg gcttgtactg aactatctag tgatcacgtt aacacctaac
5041 tctccttctt tcttccaggg gccgggctag agcgacatca tggtattccc cttactaaaa
5101 aaagtgtgta ttaggaggag agagaggaaa aaaagaggaa agaaggaaaa aaaaaagaat
5161 taaaaaaaaa aaaaagaaaa acagaagatg accttgatgg aaaaaaaata ttttttaaaa
5221 aaaagatata ctgtggaagg ggggagaatc ccataactaa ctgctgagga gggacctgct
5281 ttggggagta ggggaaggcc cagggagtgg ggcagggggc tgcttattca ctctggggat
5341 tcgccatgga cacgtctcaa ctgcgcaagc tgcttgccca tgtttccctg ccccccttcac
5401 ccccttgggc ctgctcaagg gtaggtgggc gtgggtggta ggagggtttt ttttacccag
5461 ggctctggaa ggacaccaaa ctgttctgct tgttaccttc cctcccgtct tctcctcgcc
5521 tttcacagtc ccctcctgcc tgctcctgtc cagccaggtc taccacccac cccacccctc
5581 tttctccggc tccctgcccc tccagattgc ctggtgatct attttgtttc cttttgtgtt
5641 tctttttctg ttttgagtgt ctttctttgc aggtttctgt agccggaaga tctccgttcc
5701 gctcccagcg gctccagtgt aaattccccт tccccctggg gaaatgcact accttgtttt
5761 ggggggttta ggggtgtttt tgtttttcag ttgttttgtt tttttgtttt ttttttttcc
5821 tttgcctttt ttccctttta tttggaggga atgggaggaa gtgggaacag ggaggtggga
5881 ggtggatttt gtttattttt ttagctcatt tccaggggtg ggaatttttt tttaatatgt
5941 gtcatgaata aagttgtttt tgaaaataaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
6001 aa
```

Fig. 4B